ISSN 2007-9737

# Emotional Similarity Word Embedding Model for Sentiment Analysis

Kazuyuki Matsumoto[1], Takumi Matsunaga[2], Minoru Yoshida[1], Kenji Kita[1]

[1] Tokushima University,
Graduate School of Technology, Industrial and Social Sciences,
Japan

[2] Tokushima University,
Graduate Schools of Science and Technology
for Innovation Division of Science and Technology,
Japan

{Matumoto, mino, kita} @is.tokushima-u.ac.jp, c612135014@tokushima-u.ac.jp

**Abstract.** We propose a method for constructing a dictionary of emotional expressions, which is an indispensable language resource for sentiment analysis in the Japanese. Furthermore, we propose a method for constructing a language model that reproduces emotional similarity between words, which to date has yet not been considered in conventional dictionaries and language models. In the proposed method, we pre-trained sentiment labels for the distributed representations of words. An intermediate feature vector was obtained from the pre-trained model. By learning an additional semantic label on this feature vector, we can construct an emotional semantic language model that embeds both emotion and semantics. To confirm the effectiveness of the proposed method, we conducted a simple experiment to retrieve similar emotional words using the constructed model. The results of this experiment showed that the proposed method can retrieve similar emotional words with higher accuracy than the conventional word-embedding model.

**Keywords.** Emotion recognition, emotional similarity, neural networks.

## 1 Introduction

In recent years, the distributed representation of words and sentences has been frequently used as an artificial intelligence technique to analyze data on social networking sites. Distributed representation has made it easier to calculate relevance and similarity, and to use them as features in machine learning by quantifying the features of words and sentences in the form of vectors, which were handled symbolically. However, one problem with a distributed representation of words and sentences is that it can handle semantic information, but does not deal with emotional information effectively. For example, suppose that there are two types of expressions that express emotions in a certain situation: positive and negative words. These expressions are often used in similar contexts, even if the emotions are the opposite. Many distributed expressions are based on a large corpus and are intended to extract semantic information from context, etc. If they are used as is, they are considered incapable of expressing emotions correctly, as aforementioned.

In emotion recognition, these problems do not have a significant impact because supervised machine learning is performed using distributed expressions as features. However, when generating paraphrased sentences, based on word variants, it is necessary to have a mechanism to suppress the replacement of words with those that are semantically similar but are of the opposite meaning. In addition, emotions are often analyzed not only based on polarity, such as positive/negative, but also based on basic emotions expressed in a circle of emotions, as proposed by psychologist Plutchik [1]. Ptaszynski et al. [2] conducted an emotion analysis based on

a dictionary of emotional expressions. In addition, Sano [3] created a systematic dictionary of words expressing emotions in the form of a dictionary of appraisal expressions.

The aforementioned generalized dictionary of emotional expressions, is useful not only for the emotional analysis of linguistic information, but also for facilitating communication between people. Emotional Quotient (EQ) is a measure of the intelligence required to express one's own emotions appropriately and to understand the emotions of others [4].

However, generalized dictionaries do not include unknown expressions, such as new words and popular phrases, and they have problems corresponding to the changes in language usage over time.

In this study, we focus on the strengths of unsupervised language distributed representation learning: the ability to specialize a model to a specific domain by training based on a corpus, and the ability to update the training data easily. Specifically, we convert a generalized sentiment dictionary into a numerical vector using distributed representations and pre-trained linguistic distributed representations that are specific to emotions.

The distributed representation obtained by this method may lose semantic information because it is specific to emotions. Therefore, a model based on a semantic dictionary, such as a thesaurus, is used to acquire distributed representations that contain semantic information.

This method aims to facilitate the construction of emotionally distributed representations, specialized for sentiment analysis of language as well as semantic information. To evaluate the effectiveness of the constructed model, we compared it with the conventional model of language distributed representation.

## 2 Related Works

The WordNet-Affect [5] and the Japanese Evaluative Polarity Dictionary (Kobayashi et al.) [6] are examples of the linguistic systematization of words expressing emotions. WordNet-Affect has a thesaurus of words expressing emotions; however, there is no official Japanese version.

Although some of them are translated from English to Japanese, there are many expressions that are not suitable for direct translation.

Therefore, an emotional thesaurus, specialized for the Japanese language, is needed.

The Dictionary of Emotional Expressions (Akira Nakamura) [7] is a collection of emotional expressions, in written text, from 806 works by 197 modern and contemporary Japanese authors. The dictionary defines ten types of emotion (joy, anger, sorrow, fear, shame, like, hate, excitement, relief, surprise) and compound emotions.

This dictionary contains a relatively comprehensive summary of emotional expressions used within the Japanese language.

The dictionary of appraisal expressions constructed by Sano [3], is a classification of expressions that describe values. It is unique in that it defines perspectives other than the criteria of emotion polarity, that is, positive and negative, for evaluation classification and emotion analysis. However, the dictionary does not clearly define the types of emotions; therefore, it is necessary to associate emotion classes with attributes to employ conventional emotion analysis methods.

Emo2Vec, proposed by Wang et al. [8], is based on two different models (i.e., local and global) for adding sentiment information to word vectors to analyze opinions from review sentences. This method is based on Plutchik's circle of emotions and uses multi-task learning to achieve higher expressive power than the existing emotion polarity. Their work improves on existing word and emotion embeddings adopted in experiments on the Chinese and English languages.

The difference between our method and their approach is that the emotion space is given as an 8-dimensional vector. Our study defines a 25-dimensional vector as the basic axis so that the types of emotions can be handled as flexibly as possible. To handle multiple emotions, the sigmoid function is used as the output layer to predict a 25-dimensional vector.

This allows ambiguity in phrases that correspond to multiple emotions. In addition, semantic features are pre-trained separately from word embeddings to enhance semantic expressiveness.

**Table 1.** Emotion class by Fischer

| Large class | Sub Class | Code | Class name | Example emotions |
|---|---|---|---|---|
| A | Joy | A-1-1 | Relief | looseness, peaceful, relief, solace, etc. |
| | | A-1-2 | Impression | ecstasy, delight, etc. |
| | | A-1-3 | Hope | optimistic, expectation, tympany, enthusiasm, etc. |
| | | A-1-4 | Proud | victory, boasting, adversarial quality, etc. |
| | | A-1-5 | Pleasure | contentment, feel good, briskness, etc. |
| | | A-1-6 | Excitement | ardency, alacrity, interest, gaiety, etc. |
| | | A-1-7 | Joy | ravishment, airiness, cheerfulness, etc. |
| | Love | A-2-1 | Respect | adoration, envy, beautiful, etc. |
| | | A-2-2 | Passion | desire, intoxication, adhesion, etc. |
| | | A-2-3 | Like | love, attraction, charity, chummy, etc. |
| B | Surprise | B-1-1 | Surprise | amazement, strangeness, etc. |
| C | Anger | C-1-1 | Bitter | anguish, difficult, etc. |
| | | C-1-2 | Envy | jealousy, etc. |
| | | C-1-3 | Contempt | boke, sicken, etc. |
| | | C-1-4 | Rage | umbrage, fume, etc. |
| | | C-1-5 | Scandalize | frustration, shocked, etc. |
| | | C-1-6 | Displeasure | disconcertedness, accusation, etc. |
| D | Sorrow | D-1-1 | Pity | commiseration, sympathy, empathy, etc. |
| | | D-1-2 | Alienation | isolation, loneliness, nostalgia, dejection, etc. |
| | | D-1-3 | Guilt | shame, regret, confession, abjection, etc. |
| | | D-1-4 | Disappointment | drug, fatigue, hit or miss, rejection, etc. |
| | | D-1-5 | Sorrow | despair, unluckness, dysphoria, etc. |
| | | D-1-6 | Cruel | smart, agonal, etc. |
| | Fear | D-2-1 | Anxiety | nervousness, worry, suspense, heartache, fear, etc. |
| | | D-2-2 | Warning | consternation, abasement, fret, etc. |
| E | Neutral | E-1-1 | Neutral | neutral |

## 3 Learning of Emotional Embedding

### 3.1 Emotion Class

In this study, we classified several existing dictionaries, such as the Emotional Expressions Dictionary, the Appraisal Expressions Dictionary, and the Idiom Expressions Dictionary, according to the phylogeny of emotions proposed by Fischer [9], as shown in Table 1. Based on this classification, the distributed representation of each expression was learned to extract features specific to the emotion. As the expressions registered in the dictionary of emotional expressions do not include neutral expressions, 25 classes (excluding E) are actually used for emotional classification.

### 3.2 Embedding of Word / Phrase

Traditionally, word2vec [10], fastText [11], GloVe [12], and other methods based on CBOW or skip-gram have been used for word embedding. Recently, bidirectional encoder representation from transformers (BERT) [13] has been used; this approach enables unsupervised training by considering the position of word occurrences using trans-formers and attention mechanisms. In BERT, generic unsupervised task-solving models called

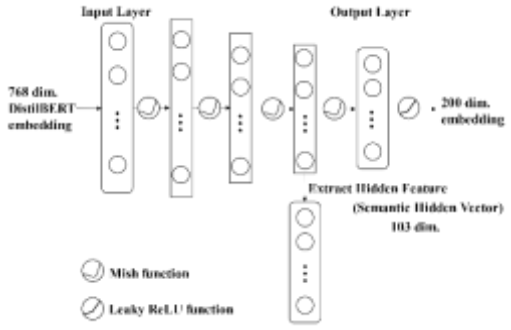


**Fig. 1.** Structure of neural networks extracting hidden semantic embedding

next sentence prediction and masked language models are pre-trained on a large unlabeled corpus.

Using the parameters obtained from the trained network, distributed representations of the language that can be applied to various tasks, are extracted. Based on the distributed representations, transfer learning or fine tuning was conducted for other tasks.

The disadvantage of BERT is that it takes a long time to train, and the trained model is large owing to the size of the network and the large number of parameters. For this reason, research has been conducted on reducing the network parameters of BERT as well as on models such as ALBERT [14] or DistilBERT (Distilled BERT) [15], which succeeded in reducing the size of the model without degrading the performance of BERT.

In this study, we target not only word units but also phrases consisting of multiple words, such as idiomatic expressions. Therefore, it is desirable to use a method that can obtain distributed expressions not only for words but also for phrases and sentences.

In our proposed method, it is necessary to convert the features obtained from words and phrases into other features that can express emotions and semantics. Therefore, we need to obtain the distributed representation to be inputted as flexibly and efficiently as possible.

Therefore, we decided to use DistilBERT, which is a lightweight model with a reduced number of parameters, as the initial embedding.

### 3.3 Sense Vector Based on Wikipedia Entity Vector

The biggest advantage of manually constructed semantic dictionaries is that they contain almost no noise, which may affect accuracy. In general, words in a manually constructed semantic dictionary belong to semantic categories, and these categories are defined by superordinate and subordinate categories.

For example, it is possible to determine the semantic distance and similarity between words, using electronic dictionaries such as WordNet [16]. However, as there are many ambiguities in language usage, it is practically impossible to construct a complete semantic dictionary that covers all the uses of a word in reality.

There have been several studies on the automatic creation of semantic dictionaries based on Wikipedia [17]. While their method can define the semantic concepts of a large number of words, they sometimes register incorrect information in the dictionary or show bias toward certain fields. However, we can extract conceptual information with high accuracy using the rich vocabulary of Wikipedia and the sophisticated information of the articles.

In this study, we consider training a model that uses the Japanese Wikipedia Entity Vector as its prediction target, which is a distributed representation of the vocabulary headings used in Wikipedia and other vocabulary used in the article, to obtain the middle-layer vector.

We used DistilBERT embeddings as inputs. Because the distributed representation vector to be predicted also contains negative values, we use Mish [18] as the activation function that can retain negative values. Figure 1 illustrates the structure of the semantic vector extraction network model. Using this model of semantic vectors together with the model of emotional vectors that will be described later, it is possible to consider both semantics and emotion. The hidden feature vector is called a semantic hidden vector (S-HV). The S-HV is a vector with 103 dimensions.

The formula for calculating Mish is shown in Equation (1), $\ln$ is the natural logarithm, and $e^x$ is the exponential function:

$$f(x) = x \cdot \tanh\left(\ln\left(1 + e^x\right)\right). \quad (1)$$

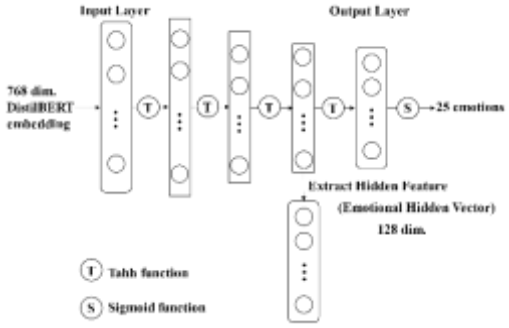


**Fig. 2.** Neural networks for learning with emotional embedding

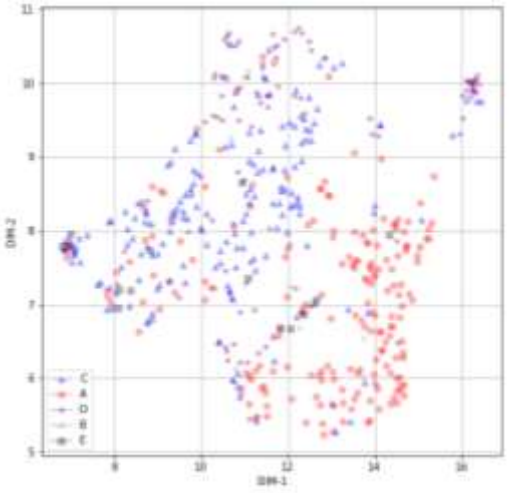


**Fig. 3.** Visualization of emotional embedding (using Neural Autoencoder and UMAP)

### 3.4 Learning Method of Emotional Embedding

To achieve embedding learning of word emotion, a model that predicts the emotion of words and phrases is required. We transformed words and phrases with emotion labels into distributed representations and then train a model to predict the emotion labels using the DistilBERT pre-training model described in Section 3.2.

The model was constructed based on a neural network. The architecture of this neural network has multiple hidden layers, and the hidden layer before the final output layer is designed to be a fully connected layer with more neurons than the number of dimensions of the emotion to be predicted.

Figure 2 shows the structure of the neural network used in this study. The hidden feature vector is called the emotional hidden vector (E-HV). The E-HV is a vector of 128 dimensions.

The results of emotional embedding, compressed using autoencoder based on neural networks, are converted to two dimensions by UMAP[19] and visualized in Fig. 3. From this figure, we can see that E (neutral) and B (surprise) are distributed in several clusters, and A (joy), C (Anger), and D (Sorrow) partially form their own clusters.

From this, it can be expected that B, which has a small number of cases, and E, which has few features, is relatively difficult to classify.

## 4 Experiment

### 4.1 Experimental Setup

We evaluated the validity of the emotional or semantic distributed representations of words, obtained by the proposed approach, using the following two methods:

**Eval-1.** We used emotion expressions with emotion labels that were not used for training as input, and calculated the similarity to the emotion expressions in the training data based on the emotion and semantic embedding of the words. From the results of this calculation, we predicted the emotion label based on the k-nearest neighbor method and obtained the correct answer rate. Experiments were conducted for the cases of k=10 and 20.

**Eval-2.** For a corpus of sentences with emotional labels, we obtained distributed representations of emotion and semantics using the trained models, and then trained sentiment prediction models using machine learning algorithms. The emotion classification model was evaluated using a cross-validation method.

The training and evaluation data for the dictionary used in Eval-1 are listed in Table 2. The

**Table 2.** Training data and evaluation data.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Train** Total: 12,180 words | | | | | | | |
| A | A-1-1 495 | A-1-2 140 | A-1-3 182 | A-1-4 851 | A-1-5 462 | A-1-6 240 | A-1-7 818 |
| | A-2-1 1,543 | A-2-2 536 | A-2-3 827 | | | | |
| B | B-1-1 784 | | | | | | |
| C | C-1-1 161 | C-1-2 47 | C-1-3 2,843 | C-1-4 1,602 | C-1-5 118 | C-1-6 758 | |
| D | D-1-1 56 | D-1-2 544 | D-1-3 124 | D-1-4 445 | D-1-5 677 | D-1-6 282 | |
| | D-2-1 590 | D-2-2 292 | | | | | |
| E | E-1-1 465 | | | | | | |
| **Test** Total: 693 words | | | | | | | |
| A | A-1-1 0 | A-1-2 16 | A-1-3 96 | A-1-4 0 | A-1-5 0 | A-1-6 40 | A-1-7 92 |
| | A-2-1 136 | A-2-2 0 | A-2-3 68 | | | | |
| B | B-1-1 0 | | | | | | |
| C | C-1-1 41 | C-1-2 0 | C-1-3 0 | C-1-4 16 | C-1-5 0 | C-1-6 96 | |
| D | D-1-1 20 | D-1-2 0 | D-1-3 24 | D-1-4 0 | D-1-5 32 | D-1-6 0 | |
| | D-2-1 40 | D-2-2 0 | | | | | |
| E | E-1-1 0 | | | | | | |

Eval-2 gradient boosting algorithm was used. LightGBM was used as the library.

For the evaluation corpus, we used Japanese sentences from the Japanese-English bilingual sentiment corpus (J-Corpus) [19, 20], and tweets and blogs with emotion tags (Web-Corpus). The tags assigned to J-Corpus were used after converting them into major categories A, B, C, D, and E.

The breakdown of the data is presented in Table 3. Table 4 shows the breakdown of words in the corpus by emotion type for both the Web-Corpus and J-Corpus, and Table 5 shows the number of words by part of speech.

We used the training data for emotion words (Train: 12,180 words) to count emotion words by emotion category. Some words, sentences, and phrases are given more than one emotion tag, because the interpretation may differ slightly from one dictionary to another.

The combinations of features to be compared are presented in Table 6. The "v" in the cells of the table indicates that the feature is used, and the "-" indicates that it is not used. To combine multiple features, each feature vector was connected horizontally.

### 4.2 Evaluation Method

In Eval-1, recall, precision, and F1-score were calculated and evaluated for each level of granularity in the hierarchy of emotion categories (1, 2, and 3 levels). The values of k were 10, 20, and 0.7, 0.5, and 0.3 were used for the similarity threshold.

**Table 3.** Evaluation corpora

| Sentence Emotion Class (Large Class) | J-Corpus | | Web-Corpus | |
|---|---|---|---|---|
| | Sentences | Words | Sentences | Words |
| A (Joy) | 212 | 2,426 | 30,777 | 436,783 |
| B (Surprise) | 22 | 306 | 1,592 | 19,973 |
| C (Anger) | 238 | 2,770 | 15,018 | 252,922 |
| D (Sorrow) | 129 | 1,587 | 21,902 | 265,259 |
| E (Neutral) | 589 | 7,106 | 7,232 | 78,513 |
| Total | 1,190 | 14,195 | 76,521 | 1,053,450 |

**Table 4.** Number of emotion words for each emotion class

| Sentence Emotion Class | J-Corpus Number of emotion words for each emotion class | | | | | Web-Corpus Number of emotion words for each emotion class | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | A | B | C | D | E |
| A | 167 | 17 | 7 | 20 | 1 | 23,698 | 1,193 | 4,326 | 1,628 | 1,047 |
| B | 7 | 7 | 12 | 8 | 1 | 512 | 260 | 212 | 134 | 288 |
| C | 69 | 13 | 152 | 38 | 1 | 6,337 | 598 | 6,036 | 2,043 | 427 |
| D | 35 | 13 | 63 | 77 | 3 | 7,310 | 749 | 4,574 | 3,546 | 586 |
| E | 197 | 46 | 253 | 141 | 18 | 2,526 | 154 | 670 | 209 | 127 |
| Total | 475 | 96 | 487 | 284 | 24 | 40,383 | 2,954 | 15,818 | 7,560 | 2,475 |

**Table 5.** Number of words for each part of speech

| Sentence Emotion Class | J-Corpus Number of words for each POS | | | Web-Corpus Number of words for each POS | | |
|---|---|---|---|---|---|---|
| | Noun | Adjective | Verb | Noun | Adjective | Verb |
| A | 785 | 79 | 281 | 149,617 | 15,969 | 55,504 |
| B | 92 | 3 | 56 | 6,826 | 497 | 2,606 |
| C | 874 | 70 | 340 | 82,468 | 6,222 | 35,874 |
| D | 495 | 36 | 203 | 85,605 | 8,974 | 37,565 |
| E | 2,169 | 151 | 1,052 | 28,646 | 1,482 | 11,353 |
| Total | 4,415 | 339 | 1,932 | 353,162 | 33,144 | 142,902 |

**Table 6.** Combination of features

| Combination ID | Combination Type | E-HV | S-HV | DBERT |
|---|---|---|---|---|
| 1 | ehv | v | - | - |
| 2 | shv | - | v | - |
| 3 | dv | - | - | v |
| 4 | ehv+shv | v | v | - |
| 5 | ehv+shv+dv | v | v | v |

In Eval-2, Recall, Precision, and F1-score were calculated for four major categories, A, B, C, and D, excluding neutral "E." 5-fold cross-validation was used to deal with class imbalance, and the Synthetic Minority Over-sampling Technique) [22], Edited Nearest Neighbor (ENN) [23], SMOTE-ENN [24], and SMOTE-Tomek Links [25] were used as resampling methods. For oversampling and undersampling, we used the class module in library imbalanced learning[1].

[1] https://imbalanced-learn.org/stable/

**Table 7.** Experimental Result of Eval-1

| Category | Comb. Type | threshold | k | Accuracy |
|---|---|---|---|---|
| Large Class | ehv+shv+dv | 0.3 | 10 | 0.595 |
| | | | 20 | 0.580 |
| | | 0.5 | 10 | 0.595 |
| | | | 20 | 0.584 |
| | **ehv+shv** | 0.7 | 10 | **0.600** |
| | ehv | | 20 | 0.590 |
| Sub Class | ehv+shv+dv | 0.3 | 10 | 0.392 |
| | **ehv** | | 20 | **0.411** |
| | ehv+shv+dv | 0.5 | 10 | 0.392 |
| | **ehv** | | 20 | **0.411** |
| | ehv+shv | 0.7 | 10 | 0.397 |
| | ehv | | 20 | 0.405 |

**Table 8.** Precision, Recall, and F1-score for each emotion class (J-Corpus)

| J-Corpus Result | | A | | | B | | | C | | | D | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Resampling** | **Comb.ID** | **Pre.** | **Rec.** | **F1** | **Pre.** | **Rec.** | **F1** | **Pre.** | **Rec.** | **F1** | **Pre.** | **Rec.** | **F1** |
| SMOTE | ehv | 0.76 | 0.76 | 0.76 | **0.38** | 0.27 | **0.32** | 0.74 | 0.77 | 0.76 | 0.53 | 0.51 | 0.52 |
| | shv | 0.56 | 0.55 | 0.56 | 0.36 | 0.18 | 0.24 | 0.58 | 0.63 | 0.6 | 0.33 | 0.32 | 0.33 |
| | dv | 0.73 | 0.73 | 0.73 | 0.15 | 0.09 | 0.11 | 0.68 | 0.77 | 0.72 | 0.53 | 0.43 | 0.48 |
| | **ehv+shv** | **0.77** | 0.76 | 0.76 | 0.33 | 0.27 | 0.3 | 0.75 | 0.82 | **0.79** | 0.6 | **0.53** | **0.56** |
| | ehv+shv+dv | 0.76 | 0.76 | 0.76 | 0.24 | 0.23 | 0.23 | 0.74 | 0.81 | 0.77 | 0.6 | 0.5 | 0.55 |
| ENN | ehv | 0.69 | 0.8 | 0.74 | 0.16 | 0.23 | 0.19 | 0.68 | 0.82 | 0.74 | 0.6 | 0.19 | 0.28 |
| | shv | 0.41 | 0.19 | 0.26 | 0.08 | 0.41 | 0.14 | 0.43 | 0.72 | 0.54 | 0 | 0 | 0 |
| | dv | 0.53 | 0.52 | 0.53 | 0.12 | 0.27 | 0.16 | 0.51 | 0.7 | 0.59 | 0.54 | 0.05 | 0.1 |
| | **ehv+shv** | 0.76 | 0.82 | **0.79** | 0.22 | 0.27 | 0.24 | 0.67 | **0.86** | 0.76 | 0.67 | 0.22 | 0.33 |
| | **ehv+shv+dv** | 0.71 | 0.79 | 0.75 | 0.07 | 0.09 | 0.08 | 0.66 | **0.86** | 0.75 | **0.71** | 0.16 | 0.25 |
| SMOTE-ENN | ehv | 0.62 | 0.86 | 0.72 | 0.16 | 0.32 | 0.21 | 0.83 | 0.5 | 0.62 | 0.5 | 0.47 | 0.48 |
| | shv | 0.4 | 0.9 | 0.55 | 0.11 | 0.36 | 0.16 | 0.58 | 0.03 | 0.06 | 0.19 | 0.05 | 0.08 |
| | **dv** | 0.45 | **0.92** | 0.6 | 0.21 | **0.55** | 0.3 | 0.77 | 0.04 | 0.08 | 0.44 | 0.33 | 0.38 |
| | ehv+shv | 0.63 | 0.9 | 0.74 | 0.19 | 0.5 | 0.28 | 0.84 | 0.47 | 0.6 | 0.48 | 0.4 | 0.44 |
| | **ehv+shv+dv** | 0.61 | 0.9 | 0.72 | 0.24 | 0.5 | **0.32** | **0.85** | 0.46 | 0.6 | 0.49 | 0.42 | 0.45 |
| SMOTE-Tomek Links | ehv | 0.74 | 0.71 | 0.73 | 0.21 | 0.27 | 0.24 | 0.75 | 0.78 | 0.76 | 0.5 | 0.47 | 0.48 |
| | shv | 0.58 | 0.57 | 0.57 | 0.24 | 0.18 | 0.21 | 0.6 | 0.63 | 0.62 | 0.36 | 0.35 | 0.35 |
| | dv | 0.73 | 0.69 | 0.71 | 0.28 | 0.23 | 0.25 | 0.66 | 0.76 | 0.71 | 0.55 | 0.47 | 0.51 |
| | **ehv+shv** | **0.77** | 0.71 | 0.74 | 0.11 | 0.14 | 0.12 | 0.72 | 0.79 | 0.75 | 0.5 | 0.47 | 0.48 |
| | **ehv+shv+dv** | 0.76 | 0.76 | 0.76 | 0.26 | 0.27 | 0.27 | 0.73 | 0.82 | 0.77 | 0.63 | 0.5 | **0.56** |

## 5 Results and Discussion

### 5.1 Result of Eval-1

In the experiment of Eval-1, only the accuracy was calculated. Table 7 shows the top similarity thresholds, k values, and feature combinations for each class hierarchy (Large, Sub).

In the large class, the combination of emotional embedding and semantic embedding has the highest accuracy.

In the sub-class, the best accuracy is obtained when only emotional embedding is used. In the emotion classification of emotional expressions,

**Table 9.** Precision, Recall, and F1-score for each emotion class (Web-Corpus)

| Web-Corpus Result | | A | | | B | | | C | | | D | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Resampling** | **Comb.ID** | **Pre.** | **Rec.** | **F1** | **Pre.** | **Rec.** | **F1** | **Pre.** | **Rec.** | **F1** | **Pre.** | **Rec.** | **F1** |
| SMOTE | ehv | 0.72 | 0.69 | 0.7 | 0.37 | 0.23 | 0.28 | 0.51 | 0.57 | 0.54 | 0.61 | 0.61 | 0.61 |
| | shv | 0.66 | 0.71 | 0.68 | 0.58 | 0.18 | 0.28 | 0.5 | 0.44 | 0.47 | 0.58 | 0.59 | 0.58 |
| | **dv** | 0.71 | 0.75 | **0.73** | **0.65** | 0.23 | **0.34** | **0.57** | 0.53 | 0.55 | 0.64 | 0.64 | **0.64** |
| | ehv+shv | 0.7 | 0.72 | 0.71 | 0.58 | 0.19 | 0.29 | 0.53 | 0.53 | 0.53 | 0.61 | 0.62 | 0.61 |
| | **ehv+shv+dv** | 0.71 | 0.74 | **0.73** | 0.63 | 0.21 | 0.31 | **0.57** | 0.54 | 0.55 | 0.63 | 0.64 | **0.64** |
| ENN | ehv | 0.54 | 0.84 | 0.66 | 0.08 | 0.06 | 0.07 | 0.46 | 0.17 | 0.25 | 0.52 | 0.35 | 0.42 |
| | **shv** | 0.49 | **0.88** | 0.63 | 0.06 | 0.04 | 0.05 | 0.32 | 0.02 | 0.05 | 0.47 | 0.25 | 0.33 |
| | dv | 0.55 | 0.81 | 0.66 | 0.09 | 0.12 | 0.1 | 0.43 | 0.14 | 0.21 | 0.52 | 0.41 | 0.46 |
| | ehv+shv | 0.55 | 0.83 | 0.66 | 0.09 | 0.07 | 0.08 | 0.45 | 0.19 | 0.26 | 0.52 | 0.37 | 0.43 |
| | ehv+shv+dv | 0.56 | 0.83 | 0.67 | 0.09 | 0.1 | 0.09 | 0.45 | 0.19 | 0.26 | 0.54 | 0.39 | 0.45 |
| SMOTE-ENN | **ehv** | 0.6 | 0.76 | 0.67 | 0.1 | **0.33** | 0.15 | 0.44 | 0.45 | 0.45 | 0.65 | 0.29 | 0.41 |
| | **shv** | 0.51 | **0.88** | 0.64 | 0.13 | 0.14 | 0.14 | 0.41 | 0.23 | 0.3 | 0.65 | 0.18 | 0.28 |
| | dv | 0.57 | 0.87 | 0.69 | 0.31 | 0.21 | 0.25 | 0.5 | 0.37 | 0.43 | 0.7 | 0.31 | 0.43 |
| | ehv+shv | 0.56 | 0.84 | 0.68 | 0.16 | 0.18 | 0.17 | 0.46 | 0.38 | 0.42 | 0.68 | 0.27 | 0.39 |
| | **ehv+shv+dv** | 0.57 | 0.87 | 0.69 | 0.28 | 0.2 | 0.23 | 0.5 | 0.38 | 0.43 | **0.71** | 0.31 | 0.43 |
| SMOTE-Tomek Links | **ehv** | **0.73** | 0.66 | 0.69 | 0.22 | 0.27 | 0.24 | 0.5 | **0.58** | 0.54 | 0.6 | 0.62 | 0.61 |
| | shv | 0.66 | 0.7 | 0.68 | 0.45 | 0.19 | 0.26 | 0.49 | 0.44 | 0.47 | 0.58 | 0.59 | 0.58 |
| | **dv** | 0.71 | 0.74 | **0.73** | 0.57 | 0.23 | 0.32 | **0.57** | 0.54 | **0.56** | 0.64 | **0.65** | **0.64** |
| | ehv+shv | 0.71 | 0.71 | 0.71 | 0.49 | 0.22 | 0.3 | 0.54 | 0.54 | 0.54 | 0.61 | 0.63 | 0.62 |
| | **ehv+shv+dv** | 0.72 | 0.74 | **0.73** | 0.6 | 0.23 | 0.33 | 0.56 | 0.54 | 0.55 | 0.63 | **0.65** | **0.64** |

emotional embedding is effective, but semantic embedding is not so effective by itself; however, if it is combined with other features, it might be effective for expressions that cannot be classified properly by other features alone.

### 5.2 Result of Eval-2

Table 8 shows the values of Precision, Recall, and F1-score for each combination of features and the resampling method when J-Corpus is used. The results showed that the feature combination (Comb. ID=5) using SMOTE (with all three types of features) yielded the best results overall. In the case where only emotional embedding (ehv) is used as a feature (Comb. ID=1), emotion B (surprise) demonstrated relatively high scores.

When semantic embedding (shv) was added to emotional embedding (Comb. ID=4), the scores for all emotions, except for emotion B (surprise), were relatively high, and the overall accuracy was also improved. This suggests that emotional and semantic embeddings can complement each other.

Figure 4 shows a graph comparing the correct answer rate, the macro-average correct answer rate, and the weighted average correct answer rate. The feature combination (Comb ID=4) (ehv+shv) exhibited the best performance. These results indicate that two features of emotional embedding and semantic embedding are effective, and SMOTE is suitable as a resampling method.

Next, the results when the Web-Corpus was used are shown in Table 9 and Figure 5, as in the case of J-Corpus. When DistilBERT was used alone, the efficiency was the highest. This may be due to the fact that, unlike Web-Corpus, Web-Corpus has many colloquial expressions, and that emoticons other than emotional expressions are used frequently in tweets and blog posts.

## 6 Conclusion

We proposed a method to learn emotional and semantic embeddings based on a Japanese dictionary of emotional expressions and using a pre-trained model as the initial feature. Because the proposed method embeds both emotions and

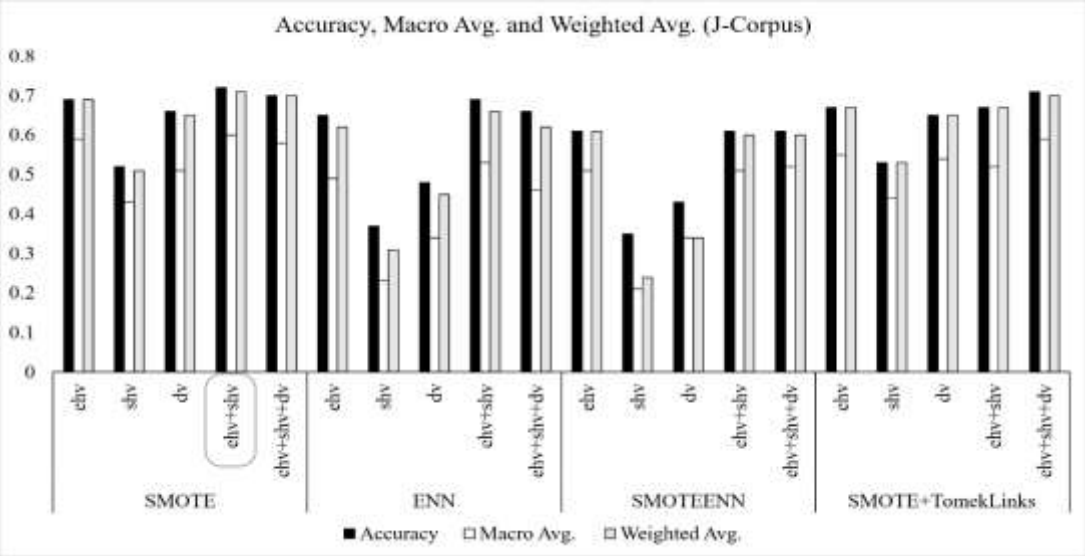


**Fig. 4.** Comparison of accuracy for each feature combination (J-Corpus)

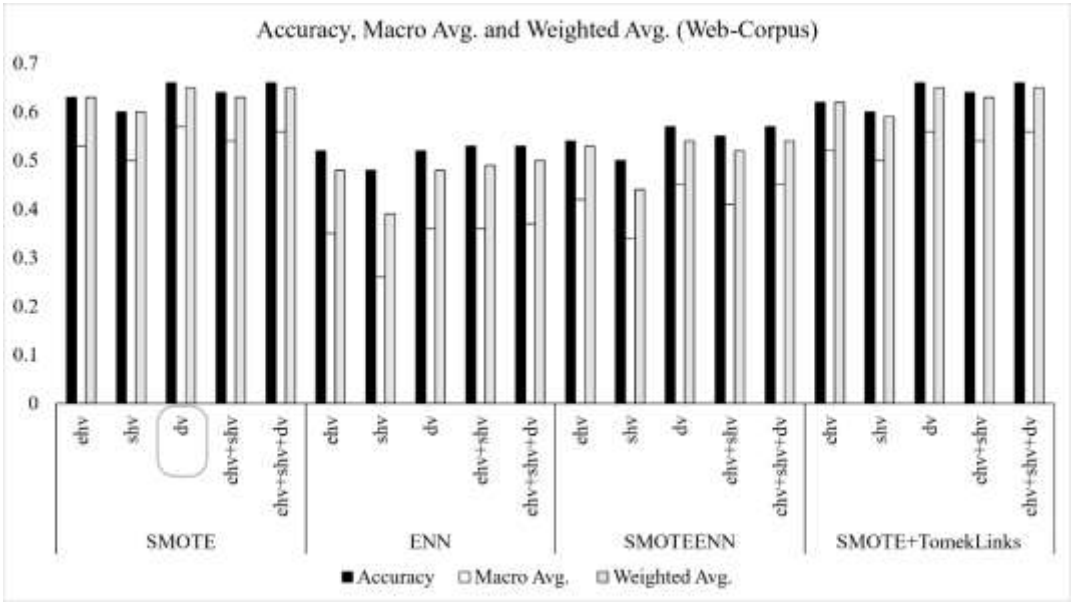


**Fig. 5.** Comparison of accuracy for each feature combination (Web-Corpus)

semantics, it can be said that it is more specialized for emotion analysis than existing language models.

To evaluate the validity of the proposed method, we conducted two experiments.

The first is a classification experiment on unknown emotional expressions based on the k-nearest neighbor method using words and phrases registered in the emotional expression dictionary.

In this experiment, using both emotional and semantic embedding, we observed a higher rate of correct answers than using only DistilBERT and demonstrated the effectiveness of the proposed method.

The other experiment was an emotion classification experiment on the corpus of utterances with the annotation of sentiment labels. We used a machine learning model based on the gradient boosting method and resampling methods, such as SMOTE, to deal with imbalances between classes, and then cross-validated the accuracy of the models.

In the experiments using the example sentence corpus, the proposed method of adding emotional embedding and semantic embedding showed better performance than using only DistilBERT's distributed representation. Meanwhile, in the experiment using the Web corpus, the performance was highest when only DistilBERT was used, indicating that it was not effective.

This may be owing to the fact that both emotion and semantic embedding are based on the data in the dictionary, and it may have been difficult to deal with the phrases unique to colloquial sentences used on the Web.

In the future, we would like to improve the accuracy by using a pre-training model that is fine-tuned based on a corpus containing a large number of colloquial sentences.

## Acknowledgments

This work was supported by JSPS KAKENHI (Grant Number JP20K12027, JP21K12141).

## References

1. **Plutchik, R. (1980).** A General Psychoevolutionary Theory of Emotion. Theories of Emotion, pp. 3–22.
2. **Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R., Araki, K. (2009).** A System for Affect Analysis of Utterances in Japanese Supported with Web Mining. Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 21, No. 2, pp. 30–49.
3. **Sano, M. (2012)**. The classification of japanese evaluative expressions and the construction of a dictionary of attitudinal lexis: An interpretation from appraisal perspective, NINJAL Research Papers, pp. 53–83.
4. **Goleman, D. (2012).** Emotional intelligence. New York: Bantam Books.
5. **Strapparava, C., Valitutti, A. (2004).** WordNet affect: An affective extension of WordNet. Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), pp. 1083–1086.
6. **Kobayashi, N., Inui, K., Matsumoto, Y., Tateishi, K. (2005).** Collecting evaluative expressions for opinion extraction. Journal of Natural Language Processing, Vol. 12, No. 3, pp. 203–222.
7. **Nakamura, A. (1993).** Kanjo hyogen jiten [Dictionary of Emotive Expression]. Tokyodo Publishing.
8. **Wang, S., Maoliniyazi, A., Wu, X., Meng, X. (2020).** Emo2Vec: Learning emotional embeddings via multi-emotion category. ACM Transactions on Internet Technology, Vol. 20, No. 2, pp. 11–17. DOI:10.1145/3372152.
9. **Fishcer, K.W. (1989).** A skill approach to emotional development: From basic- to subordinate-category emotions **Damon, W. (Ed.).** Child development today and tomorrow, pp. 107–136.
10. **Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J. (2013).** Distributed Representations of Words and Phrases and their Compositionality. Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13), Vol. 2, pp. 3111–3119.
11. **Joulin, A., Grave, E., Bojanowski, P., Mikolov, T. (2017).** Bag of Tricks for Efficient Text Classification. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Vol. 2, pp. 427–431.
12. **Pennington, J., Socher, R., Manning, C.D (2014).** GloVe: Global vectors for word representation. Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP´14), pp. 1532–1543.
13. **Devlin, J., Chang, M.W., Lee, K., Toutanova, K. (2018)**. BERT: Pre-training of deep bidirectional transformers for language understanding. http://arxiv.org/abs/1810.04805.
14. **Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R. (2020).** ALBERT: A Lite BERT for self-supervised learning of language representations. Proceedings of The International Conference on Learning Representations (ICLR2020).
15. **Sanh, V., Debut, L., Chaumond, J., Wolf, T. (2019).** DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. Thirty-third Conference on Neural Information Processing Systems (NIPS).
16. **Miller, G.A. (1995).** WordNet: A lexical database for English. Communications of the ACM, Vol. 38, No. 11, pp. 39–41.

17. **Nakayama, K., Hara, T., Nishio, S. (2006).** Wikipedia Mining to Construct a Thesaurus. Journal of Information Processing Society of Japan, Vol. 47, No. 10, pp. 2917-2928.
18. **Diganta, M. (2019).** Mish: A self regularized non-monotonic activation function. arXiv preprint arXiv:1908.08681.
19. **McInnes, L., Healy, J., Saul, N., GroBberger, L. (2018).** UMAP: Uniform Manifold Approximation and Projection. The Journal of Open-Source Software. DOI:10.21105/joss.00861.
20. **Minato, J., Matsumoto, K., Ren, F., Tsuchiya, S., Kuroiwa, S. (2008).** Evaluation of emotion estimation methods based on statistic features of emotion tagged corpus. International Journal of Innovative Computing, Information and Control, Vol. 4, No. 8, pp. 1931–1941.
21. **Matsumoto, K., Ren, F. (2011).** Estimation of word emotions based on part of speech and positional information. Computers in Human Behavior, Vol. 27, No. 5, pp. 1553–1564.
22. **Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002).** SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, Vol. 16, pp. 321–357.
23. **Wilson, D. (1972).** Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 2, No. 3, pp. 408-421.
24. **Batista, G., Bazzan, A., Monard, M. (2003).** Balancing training data for automated annotation of keywords: A case study. Proceedings of the 2nd Brazilian Workshop on Bioinformatics, pp. 10–18.
25. **Batista, G., Prati, R., Monard, M. (2004).** A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter, Vol. 6, No. 1, pp. 20–29.

*Article received on 14/06/2021; accepted 05/09/2021.*
*Corresponding author is Kazuyuki Matsumoto.*